REMARKS/ARGUMENTS

Favorable reconsideration of this application, as presently amended and in light of the following discussion, is respectfully requested.

Claims 29-50 and 54-58 are pending in this application, of which Claims 29 and 49 are amended, and Claims 57-58 are new. Support for the changes to the claims is found in the originally filed disclosure, including the specification at least from page 27, line 31 to page 28, line 19. No new matter is added.

In the outstanding Office Action, Claims 29, 32-39, 43, 46-50 and 54-56 were rejected under 35 U.S.C. § 103(a) as unpatentable over U.S. 6,593,956 (Potts) in view of U.S. 2003/0030735 (Ike) and U.S. 2003/0085997 (Takagi); Claims 30 and 31 were rejected under 35 U.S.C. § 103(a) as unpatentable over Potts in view of Ike, Takagi and U.S. 6,408,301 (Patton); Claims 40-42 were rejected under 35 U.S.C. § 103(a) as unpatentable over Potts in view of Ike, Takagi and U.S. 6,297,846 (Edanami); and Claims 44 and 45 were rejected under 35 U.S.C. § 103(a) as unpatentable over Potts in view of Ike, Takagi and US 2003/0035479 (Kan).

An aspect of this application is to provide a system to detect faces of different sizes in an image, where the image may be scaled by a range of factors, where a distance (i.e., probability) map is produced for each scale.[1] Figures 13A-13C of the application show images and corresponding distance maps for three different scales. This is performed to provide a relatively highest probability among all of the probability maps at all of the scales.[2] However, searching for faces at multiple scales adds additional computational cost and also the potential for a detection errors.

Turning now to the claim amendments submitted herewith, to assist with the above-noted process, the claims define the use of focus and zoom data to determine a distance of a

[1] Specification, page 11, lines 27-28.
[2] Specification, page 12, lines 6-8.

face from the video camera, and thus an expected face size. The lens focus and zoom setting is also used to, based on the expected face size, either calculate a subset of image scales for face detection within the captured images or to calculate weighting factors for the image scales to variably weight the image scales for face detection within the captured images. See, for example, Claims 29 and 57.

In other words, consistent with these claims, the focus and zoom settings of a camera can give an initial indication of an expected image size of a face that may be present in the image. Thus, with reference to (for example) an average face size, it is possible to calculate an expected size of a face. This expected face size can be a pixel measurement in image data, as described in a non-limiting example in the specification.[3] This expected face size can lead to a subset of scales for searching or a variable weighting, as is discussed above.

Although varying in scope and/or directed to different statutory classes, Claims 49 and 58 recite features which are substantially similar to those noted above in Claims 29 and 57. It is respectfully submitted the cited references fail to disclose or reasonably suggest the features defined by these claims.

In particular, Figure 4 of Potts shows that faces are detected within a video as a first step 102. This is discussed briefly at column 7, lines 59-61 of Potts, but is discussed in more detail with reference to Figure 5. Figure 4 and column 8, lines 50-59 of Potts states that an audio range data and video coordinates are used to pan, tilt and zoom the camera on a current speaker.

Thus, it appears clear from Potts that there is a default initial setting in which faces are detected and located in the video image prior to any change in zoom, and this data is used in conjunction with audio-based detection of who is talking and their range in order to then zoom in on the current speaker.

[3] Specification, page 28, lines 17-19.

The text accompanying Figure 5 of Potts (namely, column 8, line 64 to column 10, line 62) makes clear that the facial detection system is quite simple, and relies on motion in skin-colored pixels. This suggests at column 10, lines 40-48 that candidate faces are rejected if they are not a "default size" at the camera range value. Therefore, Potts teaches that before any zoom is applied, candidate faces that are not of a default size are rejected in an image at a default "camera range value" which is a default setting prior to application of any zoom. In Potts, the zoom is only used to single out a particular speaker that has already been detected in the video image.[4] The zoom is controlled by the audio range finder. However, when a face is the "wrong size," then the zoom is adjusted so the face ends up being of the predetermined size.

In light of the above, Potts merely describes two modes. The first mode is a pre-zoom mode where faces are of a "default size" and rejected when they are not the default size. The second mode is a zoom mode where again faces are expected to be of a predetermined face size (***because the zoom is supposed to be based on the range of the current speaker and hence always frame the speaker properly***), but when the face in the zoomed image is not of the predetermined size, the zoom is adjusted until it is.

Consequently, the modes of Potts are clearly distinct from the claimed invention, because Potts does not disclose or reasonably suggest the face detector configured to detect faces at different image scales, where the face detector receives a lens focus and a zoom setting to determine a distance of a face from the video camera to calculate an expected face size. Further, Potts is silent regarding, based on the expected face size, either calculating face detection weighting factors or calculating a subset of the scales to search within for face detection. As a result, it is respectfully submitted the claims are allowable over Potts.

[4] Potts, column 21, line 42 to column 22, line 5.

Prior to addressing the other cited references, it should be appreciated that Potts uses an audio range finder to initially control zoom, followed by an adjustment to fit a predetermined face size. Thus, there is no mention whatsoever of focus and there is no apparent need for focus to be used with the face detection system of Potts. Consequently, any combination of Potts with a system that uses focal data clearly relies upon impermissible hindsight and is selected only and specifically to account for the features of the presently-claimed invention. In other words, there is no reason one of ordinary skill in the art would modify Potts with a system that uses focal data absent impermissible hindsight.

Nonetheless, it is respectfully submitted the other cited references fail to remedy any of the above-noted deficiencies of Potts.

The Office Action at page 4 states Ike teaches a lens focus and a zoom setting are both used to determine a distance of a face from a video camera.[5] However, Ike is entirely silent as to the word "face." Ike is silent regarding any form of face or object detection.

Ike describes a fixed position security camera that implements pre-program moves to cover different viewpoints at predetermined distances, such as doors and windows.[6] The system in Ike uses a complex pre-calculated relation (tracking curve) between zoom and focus to maintain focus from a current position to a new zoom position.[7]

In Ike, the zoom and focus for a prescribed target are stored in memory and are used to navigate a tracking curve from the current position to the target position so that the target remains in focus throughout the zoom process.[8] However, the distance of the object is calculated from ***stored*** zoom in focus data.[9] There is no discussion in Ike relating to calculating an object size (or an expected face size as required by the claims).

---

[5] Office Action, page 4, citing paragraphs [0045]-[0046] of Ike.
[6] Ike, paragraph [0001], [0027], [0032], and [0038]-[0039].
[7] Ike, paragraph [0042].
[8] Ike, Figure 5 and paragraphs [0039], [0045]-[0046].
[9] Ike, paragraph [0039].

Moreover, at no point does Ike describe use of a current, instantaneous set of focus and zoom data to determine the likely size of a face, much less the use of that information to effect the processing of a multi-scale facial detection system. If one were to combine the teachings of Potts and Ike, the only technically feasible result that does not use hindsight is that the audio range data from Potts is used to determine the required zoom in focus to frame a particular talker, and the system of Ike zooms in on that talker while maintaining focus during the zoom transition. The combination does not read on the amended claims.

None of the other cited references address these deficiencies of Potts and Ike. Accordingly, it is respectfully submitted the claims are allowable over the art of record, and the outstanding rejection should be withdrawn.

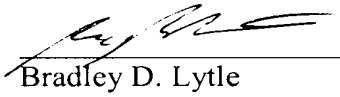
Should the Examiner disagree, the Examiner is encouraged to contact the undersigned to discuss any of the above issues. Otherwise, it is respectfully submitted no issues remain pending in this application and this application is in condition for allowance. Therefore, a timely Notice of Allowance is respectfully requested.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, L.L.P.

Customer Number
22850

Tel: (703) 413-3000
Fax: (703) 413-2220
(OSMMN 07/09)



Bradley D. Lytle
Attorney of Record
Registration No. 40,073

Marc A. Robinson
Registration No. 59,276